DOCUMENT RESUME

ED 474 625 CS 511 905

AUTHOR	Wixson, Karen K.; Fisk, Maria Chesley; Dutro, Elizabeth; McDaniel, Julie
TITLE	The Alignment of State Standards and Assessments in Elementary Reading. CIERA Report.
INSTITUTION	Center for the Improvement of Early Reading Achievement, Ann Arbor, MI.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CIERA-R-3-024
PUB DATE	2002-10-03
NOTE	53p.; Commissioned by the National Research Council's Committee on Title I Testing and Assessment.
CONTRACT	R305R70004
AVAILABLE FROM	CIERA/University of Michigan, 610 E. University Ave., 1600 SEB, Ann Arbor, MI 48109-1259. Tel: 734-647-6940; Fax: 734-763-1229; Web site: http://www.ciera.org. For full text: http://www.ciera.org/library/reports/inquiry-3/3-024/ 3-024.pdf.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	Case Studies; *Educational Assessment; Elementary Education; National Surveys; *Reading Achievement; *State Standards; Testing
IDENTIFIERS	National Research Council; Title I Evaluation and Reporting System

ABSTRACT

A study examined the alignment between state standards and assessments in elementary reading. The impetus for the study was a request by the National Research Council's Committee on Title I Testing and Assessment for information on the extent to which students' performance on state assessments could be assumed to provide evidence of their level of achievement of state standards. Concerns about alignment arose from the context surrounding Title I legislation. The sample for the survey of the alignment of state standards and assessments in elementary reading comprised all 50 states and the District of Columbia. Data were collected in late 1998 and early 1999. Criteria for evaluation originally developed for math and science were adapted for this study. Results were categorized according to current status, characteristics of standards and assessments program, and alignment. Several state case studies are presented. Findings suggest that the issue of alignment between state standards and assessment in elementary reading is more complex than it appears at first glance and that it means different things in different states. Appended are the interview protocol, sample coding sheet, cognitive levels rubric, and structure of knowledge rubric. (Contains 1 figure, 20 tables, and 6 references.) (NKA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 474 625

The Alignment of State Standards and Assessments in Elementary Reading

Karen K. Wixson
Maria Chesley Fisk
Elizabeth Dutro
Julie McDaniel
The University of Michigan

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CIERA REPORT #3-024
October 3, 2002

CIERA
Center for the Improvement of Early Reading Achievement

BEST COPY AVAILABLE

CS 511 905

The Alignment of State Standards and Assessments in Elementary Reading

CIERA REPORT #3-024

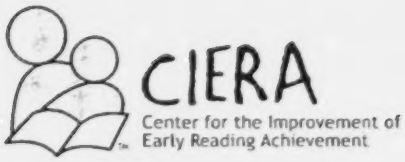
Karen K. Wixson, Maria Chesley Fisk,
Elizabeth Dutro, and Julie McDaniel
The University of Michigan

CIERA Inquiry 3: Policy and Profession
Is students' performance on state assessments evidence of their level of achievement of state standards?

This report examines the alignment between state standards and assessments in elementary reading. The impetus for our study was a request by the National Research Council's Committee on Title I Testing and Assessment for information on the extent to which students' performance on state assessments could be assumed to provide evidence of their level of achievement of state standards. We begin our report by providing the context surrounding the Title 1 legislation, which gave rise to concerns about alignment.

Center for the Improvement of Early Reading Achievement
University of Michigan – Ann Arbor

October 3, 2002



University of Michigan School of Education

610 E University Ave., Rm 1600 SEB
Ann Arbor, MI 48109-1259

734.647.6940 voice
734.615.4858 fax
ciera@umich.edu

www.ciera.org

This report was commissioned by the National Research Council's Committee on Title I Testing and Assessment

©2002 Center for the Improvement of Early Reading Achievement.

This research was supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. However, the comments do not necessarily represent the positions or policies of the National Institute of Student Achievement, Curriculum, and Assessment or the National Institute on Early Childhood Development, or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

THE ALIGNMENT OF STATE STANDARDS AND ASSESSMENTS IN ELEMENTARY READING

Karen K. Wixson
Maria Chesley Fisk
Elizabeth Dutro
Julie McDaniel

I. The Alignment of State Standards and Assessments in Elementary Reading

This report examines the alignment between state standards and assessments in elementary reading. The impetus for our study was a request by the National Research Council's Committee on Title I Testing and Assessment for information on the extent to which students' performance on state assessments could be assumed to provide evidence of their level of achievement of state standards. We begin our report by providing the context surrounding the Title 1 legislation, which gave rise to concerns about alignment.

The 1994 Improving America's Schools Act (IASA) reauthorized Chapter 1 of the Elementary and Secondary Education Act (and returned Chapter 1 to its original name, Title I). This reauthorization brought with it some dramatically different strategies for meeting the educational needs of disadvantaged children. The new Title I calls for high standards for all children, and systemic reform strategies to enable all children to achieve these standards. Specifically, the Title I statute states that the standards:

> shall include: (i) challenging content standards in academic subjects that (I) specify what students are expected to know and be able to do; (II) contain coherent and rigorous content; and (III) encourage the teaching of advanced skills; (ii) challenging student performance standards that (I) are aligned with the state's content standards; (II) describe two levels of performance, proficient and advanced, that determine how well children are mastering the material in the state content standards; and (III) describe a lower level of performance, practically proficient, to provide complete information about the progress of lower performing children toward achieving to the proficient and advanced levels of performance. For the subjects in which students will be served under this part,... the state plan shall describe a strategy for ensuring that

> [Title I] students are taught the same knowledge and skills and are held to the same expectations as are all children (as quoted in Rothman & Elmore, 1997).

The Title I statute makes the link between standards and assessments apparent by requiring states to develop assessments that are "aligned with the state's challenging content and performance standards and provide coherent information about student attainment of such standards" (IASA, U.S. Congress, 1994, p. 8). Assessments must be aligned to standards; otherwise, students preparing to do well on the tests will be performing tasks unrelated to the standards, and parents and the community will receive misleading information about children's performance (Rothman & Elmore, 1997). Compliance with this legislation has meant that states have new, critical roles to play. The high-quality, challenging standards and assessments at the core of this legislation are to be created by the states, not mandated by the federal government. States must develop their own content and performance standards and high-quality, carefully aligned assessments in order to determine how well children are meeting those standards (Payzant & Levin, 1993, p. 70).

State' progress in meeting the law's requirements was the subject of a report by Mary Jean LeTendre, the U.S. Department of Education's Director of Title I, to the NRC Committee on December 5, 1997 (Committee minutes). At the time of this report, Dr. LeTendre indicated that the Department of Education had approved the content and performance standards for 18 states; 19 states had received waivers and had until May 1998 to submit evidence that their standards were acceptable; another 4 states had waivers pending, and 9 states could not foresee meeting the May deadline. She also reported that most states had developed content standards, but many appeared to be having problems coming up with performance standards. The problem was due in part to the fact that states were not developing performance assessments on the basis of their content standards—rather, they were using existing tests that may not have been completely aligned with those standards.

At the same December, 1997, meeting, Ed Roeber, then of the Council of Chief State School Officers, reported that states had widely varied systems of assessment in place. He indicated that this situation came about in part because test publishers had been successful in convincing state officials that their tests measured state standards, when in fact the tests measured students' understanding of what was being taught. Studies that have been conducted show little correlation between tests and standards (e.g., Smith, 1997).

The importance of alignment to systemic reform in general, and Title I reform in particular, means that determining the extent of alignment within a system is a critical step in evaluating the success of reform efforts. According to Webb (1997), two or more of a system's components are aligned if they are in agreement or match each other. In educational discourse, the concept of alignment most commonly refers to the match between an assessment instrument (or instruments) and a curriculum. Both expectations and assessments are now of great concern to educators and policymakers, as the keys to standards-based education, systemic reform, and accountability.

Because of the centrality of expectations and assessments to current thinking on reform, the Committee was particularly interested in the status of state alignment of these two elements in the core Title I subject areas of mathematics and reading. Norman Webb had already undertaken an evaluation of alignment in mathematics and science that could provide information about mathematics, and the study reported here focused on elementary reading. Because the time available for this study was short and the Committee was interested in comparisons between mathematics and reading, We adapted the general concepts and procedures concerning alignment which were developed for mathematics (Webb, 1997) .

Webb (1997) defines alignment as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide students' learning toward what they are expected to know and do (p. 4). As such, alignment is considered a quality of the relationship between expectations and assessments, and not an attribute of one of these elements independent of the other. Alignment is not limited to a comparison between a single assessment instrument and a curriculum, but extends to a set of assessment instruments or the assessment system as a whole. Webb begins by quoting a report of the Mathematical Sciences Education Board (MSEB):

> The term alignment is often used to characterize the congruence that must exist between an assessment and the curriculum. Alignment should be looked at over time and across instruments" (MSEB, 1993, p. 123). A single assessment may not be well aligned with curriculum because it is too narrowly focused, but it may be part of a more comprehensive collection of assessments that is in full alignment with the curriculum. (Webb, 1997, p. 3)

It is difficult to judge alignment for several reasons. Expectations and assessments are frequently expressed in multiple documents, making it difficult to assemble a complete picture. It is also difficult to establish a common language for describing different policy elements. Finally, the policy environment in an educational system can be constantly changing. For example, new goals are sometimes mandated while old forms of assessment are still in place.

The most common methods used by states to align components of their educational systems are described by Webb (1997) as: sequential development; expert review; and document analysis. The sequential development approach involves the development of standards, frameworks, and assessments in sequence, so that each component is aligned with the one from which it is derived. Alignment by sequential development is frequently controlled within an agency, and is less likely than the other types to include form of external review. In the absence of such a review, alignment can be strengthened by incorporating checking procedures that can be used by agency staff. One disadvantage of the sequential development approach is the amount of time needed to put a program in place. This approach also ignores the need for synergy among policy elements. Even when states declare that they have established alignment through sequential development, the process itself is actually much more dynamic and recursive than linear.

The expert review approach involves convening a panel of experts to review system components and judge the quality and extent of their align-

ment. Some states have built external review panels into the process for developing important elements of their system. The quality of expert reviews depends in part on the qualifications and expertise of the reviewers. Content area specialists are essential for any review panel which is judging the match between expectations and assessment. Providing opportunities for reviewers to interact and build consensus helps improve the quality of the review.

The document analysis approach involves coding and analyzing the match among documents that convey expectations and assessments. This is the approach undertaken by Webb and the authors of the present report, as well as by other alignment studies external to the states, such as ACHIEVE and TIMSS. The document analysis approach requires use of a common metric to compare the curriculum and assessments. The reliability of the partitioning and coding of documents can be checked using sampling techniques.

The fall, 1995, survey by the Chief State School Officers, reported in Webb (1997), suggests that the sequential development method may be the approach most frequently used by states . In this survey, state assessment directors were asked, "What does alignment mean in your state?" The most common response was that assessment activities and content standards were aligned by design. For example, "aligned means assessments will be based on the standards and indicators" or "the assessments are actually... designed to measure... outcomes and requirements stated in goals and objectives. Committees approve and reject items based upon their fit with goals and objectives." Or "Curriculum frameworks provide the assessment framework for developing tests. All test questions, etc. are developed to meet the curriculum objectives." In most of the states, frameworks and assessments were judged to be aligned if goals and learning objectives were considered in some way in the design or selection of the assessment instruments (or vice versa). Most states lacked a formal, systematic process for determining the alignment among standards, frameworks, and assessments.

Our document analysis of alignment in elementary reading was guided by the following two questions, among others: How can we characterize the alignment of state standards and assessments in elementary reading? How can we characterize the document analysis method used to evaluate alignment in elementary reading?

II. Method

Sample

Our sample for this survey of the alignment of state standards and assessments in elementary reading comprised all 50 states and the District of Columbia . The data were collected and analyzed in late 1998 and early 1999.

A sample of four states was selected for more in-depth analysis, from the pool of states with approved standards and assessments that the states themselves had indicated were in alignment. This sample was selected pursuant to the Committee's advice that the study needed to provide more in-depth analyses of states with a variety of approaches to, and histories of, reform in general, and alignment in particular. The four selected states can be characterized as follows. State A has a set of mastery-oriented standards and uses a commercially published, norm-referenced test as its assessment. State B has a long history of promoting a heavily skills-oriented curriculum, and uses a state-developed, objective-referenced exam. State C is fairly new to reform and uses a combination of a norm-referenced and a state-developed exam. State D has a long history as a reforming state and uses a nationally developed criterion-referenced exam, in combination with an individualized oral reading assessment.

Procedures

We began our investigation with the intention of doing a fairly in-depth analysis of a small number of states. Our first step, was to identify the pool of states which had indicated that their approved reading standards and assessments were in alignment. To identify this pool, we gathered information about the status of reading standards and assessments from all 50 states and the District of Columbia through brief telephone interviews. This preliminary information revealed some interesting trends, so we decided to expand our telephone interviews and include information from the 50 states and the District of Columbia as part of our report. A list of the interview questions used to gather information is provided in Appendix A.

From the information gathered in the initial interviews, we identified a pool of states from which we could select a sample for more in-depth analyses. All states which reported the alignment of approved standards and assessments were included in this pool. With the advice of the Committee, we further narrowed the pool and ultimately selected four states that represented different approaches to alignment. The National Academy of Sciences and the National Reading Council mailed out letters describing the study and asking the selected states to participate. Nondisclosure agreements were signed and we were provided with the assessments needed to conduct our analyses.

Criteria for Evaluation

We adapted criteria which were originally developed for mathematics and science, and used them to evaluate the alignment of reading standards and assessments in our four sample states. These mathematics and science criteria were developed with the input of an expert panel formed by the National Institute for Science Education (NISE) and the Council of Chief State School Officers (CCSSO) (Webb, 1997). These criteria focused on five areas: Content; Articulation across Grades and Ages; Equity and Fairness; Pedagogical Implications; and System Applicability. Given the purposes of our

investigation and the short amount of time available for completion of the study, only the Content criteria were used for the analyses reported here.

The mathematics and science criteria were intended to provide a means for thinking about alignment. As such, they refer to the correspondence or comparability between standards and assessments. The Content criteria (Webb, 1997) include six subcategories:

1. Categorical Concurrence: The same categories of content, such as subject headings and their subheadings, appear in each.

2. Depth of Knowledge Consistency: The level of cognitive complexity of information students should be expected to know and demonstrate.

3. Range of Knowledge Correspondence: Standards and assessments address a comparable span of topics and ideas.

4. Structure of Knowledge Comparability: The underlying concepts of the discipline and what it means to "know" these concepts are in agreement.

5. Balance of Representation: Similar emphasis is given to different content topics, instructional activities, and tasks.

6. Dispositional Consonance: The extent to which both standards and assessments include elements such as attitudes and beliefs that go beyond learning concepts, procedures, and their applications.

In working with these criteria, we found that Categorical Concurrence was not a good indicator of alignment in reading, due largely to the variety of ways in which states have chosen to deal with reading and the other language arts in their standards and assessments. We also dropped the Dispositional Consonance category when we added a new criterion, which we call Coverage. The Coverage criterion addresses the extent to which the objectives, both within each standard and overall, are represented by at least one assessment item. This analysis allowed us to determine whether there were dispositional standards/objectives that were not addressed by the assessment—thus eliminating the need for a separate Dispositional Consonance category.

As a result of these changes, we ultimately evaluated the alignment of standards and benchmarks in terms of five criteria: Range of Knowledge Correspondence; Balance of Representation; Coverage; Depth of Knowledge Consistency; and Structure of Knowledge Comparability. It was also necessary to determine exactly how to identify standards and assessments that were specific to reading. We addressed this issue by analyzing all standards and objectives that specified reading in their titles. In cases where standards and objectives were integrated across the language arts and reading was not identified separately, we analyzed the entire set of standards and objectives. In analyzing the assessments, we eliminated from consideration any assessment items that were not directly tied to a reading score of some type.

Coding and Analysis

The standards and assessments for each of the four focus states were coded in the following manner. Each standard and its subordinate objectives were typed down one side of a spread sheet and rated for cognitive complexity. Assessment items were then numbered, and the numbers were placed across the top of this spread sheet. The coding of each assessment item according to the cognitive complexity rubric was then put into every cell which corresponded to an objective that the item appeared to measure. Each assessment item, therefore, might be related to more than one standard or objective. Assessment items that did not "hit" any objectives were simply included in the tally of the total number of assessment items. A sample coding sheet is presented in Appendix B.

The analyses conducted to examine each of the criteria were as follows:

- Range of Knowledge Correspondence and Balance of Representation—These two criteria were evaluated by calculating the number and proportion of the total number of objectives that were related to each standard, and comparing them with the number and proportion of the total number of assessment items that were related to each standard.

- Coverage—This criterion was evaluated by calculating the number and proportion of objectives under each standard that were represented by at least one assessment item, and comparing them to the total number of objectives represented by at least one assessment item.

- Depth of Knowledge Consistency—This criterion was evaluated by comparing the number and proportion of objectives within each standard that was rated at each of the four cognitive levels with the number and proportion of test items which were related to each standard that was rated at each of the four cognitive levels. Comparisons were also made between the total number of objectives rated at each of the four cognitive levels and the total number of assessment items rated at each cognitive level. Finally, a comparison was made between the overall average rating of the cognitive level of the objectives and the overall average rating of the cognitive level of the items. In order to evaluate Depth of Knowledge Consistency it was necessary to modify the cognitive complexity rubric developed for mathematics and science. One important difference between the reading and mathematics/science rubrics was the need for the reading rubric to evaluate assessment items in relation to the stimulus text to which it referred. An item that on its own may have appeared to be fairly high-level could turn out to be fairly literal when one considered the text to which it referred. The rubric is presented in Appendix C.

- Structure of Knowledge Comparability—This criterion refers to the philosophy underlying the standards and assessments. Standards and assessments were evaluated in terms of the extent to which they appeared to emanate from one of three conceptual models of reading curriculum and instruction: mastery, cognitive, or social-constructivist. See Appendix D for the descriptions of the models used for this evaluation.

We "trained" ourselves in this process by analyzing the alignment of standards and assessments from one of the states in our sample of four. As we did

so, we developed and refined a list of decision rules that guided our evaluations of which objectives were "hit," or assessed, by each item. The resulting list of decision rules reads as follows:

1. An item that hits part of an objective is considered a hit even if it does not hit all of the objective. (For example, hit "extend literal meaning of text by making inferences, and evaluate the significance and validity of texts in light of prior knowledge and experience" if item only gets at extending the literal meaning of the text by making inferences.)

2. If an item hits a standard, but not an objective within that standard, hit the standard. The standard is then counted as an objective in the analysis.

3. Some standards do not appear to encompass all of the objectives that appear under them. Hit the objective in this case. (For example, some may feel that "identify and summarize main ideas and key points from literature, informational texts, nonprint..." does not fit under the standard "read and discuss literary and nonliterary texts in order to understand human experience." In this case, hit the objective " identify and summarize main ideas and key points from literature, informational texts, nonprint..." instead of the standard.)

4. When an item requires an accumulation of skills—for example, comprehending, recalling, and interpreting—hit all objectives that correspond to each skill.

After working together on the evaluation of one state and refining both the rubric for cognitive levels and the decision rules, we proceeded to the next state. Following Webb's recommendations for training procedures, we calibrated our analyses by discussing a few "anchor," or typical, items. After some slight adjustments following the calibration, we analyzed our independent ratings. We agreed on the cognitive levels of 80% of the items and 94% of the objectives. Ninety-four percent of our "hits," or evaluations of the objectives assessed by each item, were the same. We decided that this level of agreement was sufficient for us to independently analyze the final two states.

III. Survey Results

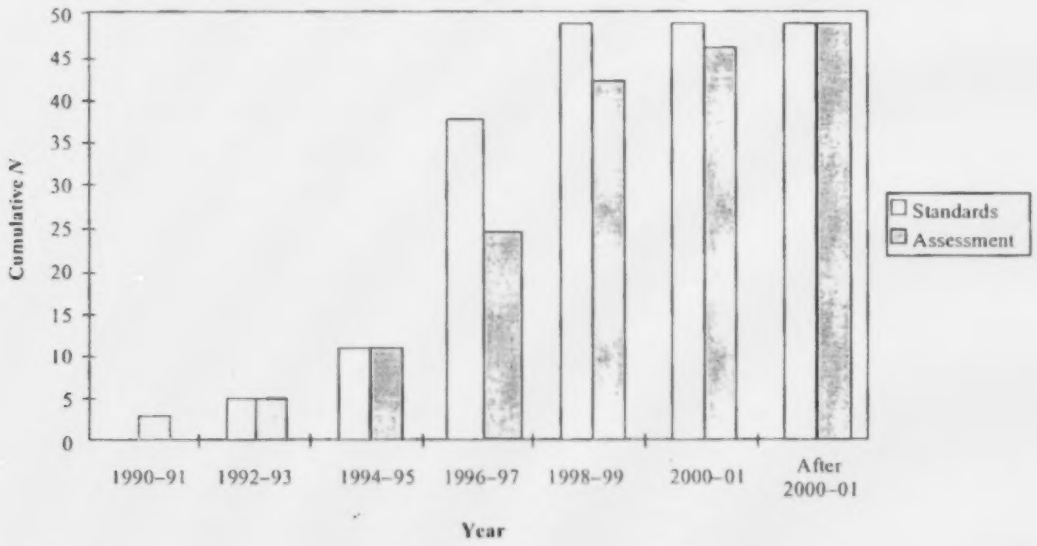
The results of the state survey are presented here in three parts: Current Status; Characteristics of Standards and Assessment Programs; and Alignment.

Current Status

Our survey of statewide standards and assessments revealed that virtually all states have actively moved toward establishing standards and assessments (see Figure 1). Specifically, 50 states (98%) currently have or will have in place statewide standards by the end of 1999. The lone exception is a state that requires local districts to develop and implement their own standards.

This state does not plan to give a statewide assessment. At least 40 states (78%) will have implemented statewide assessment systems to go along with their standards by the end of 1999. Three states reported that the development of their systems was in progress and they did not have a scheduled implementation date. It seems reasonable to expect that by the end of the 2000–2001 school year all but one (98%) state would have put their assessment programs into place.

Figure 1: Cumulative Frequencies of Years State Standards Were Approved and Assessments Implemented



*Standards for one state are determined locally. Information for one state assessment was not available.

The increase in the number of approved state standards from 1994 to 1996 appears to correspond to changes in Title I requirements. The period from 1996 to 1999 reflects the greatest increase in the number of statewide assessment systems being used for the first time, suggesting that in most cases assessments have followed close behind standards approval.

Characteristics of Standards and Assessments

Our predictions about the variety of ways in which the domain of English language arts is conceptualized and parsed were most evident in the categories used for reporting student performance (see Table 1). There were few similarities among states in these categories, with the exception of those reporting general reading (32 states, or 64%), writing (27 states, or 54%) or language arts (17 states, or 34%) scores. Some states break general reading, writing, and language arts scores down into subcategories such as vocabulary ability or language mechanics. Others include categories distinct to other areas of the language arts, such as speaking, listening, and viewing.

BEST COPY AVAILABLE

Table 1: Categories for Reporting Reading/Language Arts Scores

Category	Percentage (*N*)
General Reading Score	64% (32)
Reading Comprehension	24% (12)
Specific Reading Components (including constructing meaning, critical stance, appreciation and respect, analytical ability, reading strategies, and/or extending meaning)	18% (9)
Reading and Listening	2% (1)
Listening	16% (8)
Vocabulary	18% (9)
Spelling	12% (6)
Language Usage (including mechanics, punctuation, and/or grammar)	16% (8)
General Language Arts	34% (17)
General Writing Score	54% (27)
Writing Components Score (including organization, voice, effectiveness, editing, and/or conventions)	8% (4)
Speaking and Writing	2% (1)
Speaking	6% (3)
Discussion	2% (1)
Viewing	4% (2)
By Reading/Language Arts Standard	2% (1)
Functional Literacy	2% (1)
Not Available	2% (1)

With particular attention to the standards themselves, we found that the largest number of states—36 (72%)—cluster their standards around the primary/elementary, intermediate/middle school, and high school levels. Twelve other states (24%) have developed standards for each grade level. Two of these 12 states have also developed standards for each course taken at the high school level. In contrast, 2 states (4%) have written K-12 standards that allow school districts to meet local student needs as they deem best.

The assessments also show a predominant focus on the primary, intermediate, and high school levels (see Table 2). At present, only one state (2%) assesses students at every grade. The most frequently assessed grade in reading/language arts is the eighth grade, with 44 states (88%) testing students at that level. Next on the list is the fourth grade, with 39 states (78%) assessing students at this grade. It appears that greater emphasis is given to primary and intermediate levels, specifically the elementary school level, with 52% of the states assessing students in the third grade, and 56% assessing them in the fifth grade.

Table 2: Grades Assessed (All States Reporting)

Grades	Percentage (N = 50)
K-2	6% (3)
3	52% (26)
4	78% (39)
5	56% (28)
6	40% (20)
7	32% (16)
8	88% (44)
9	24% (12)
10	40% (20)
11	48% (24)
12	12% (6)
Every grade	2% (1)
HS in development	2% (1)
In development	2% (1)

Our accounting of states' elementary reading assessments shows a great variety in the types of reading assessments they use (see Table 3). State-developed assessments are the most widely used: 31 (60%) of the states have developed their own assessments for measuring reading abilities. The second most frequent type of assessment is the commercial, "off-the-shelf" type, which is used by 26 (51%) of the states. Three states (6%) allow districts to choose among a selection of commercially available tests. Four states (8%) use a custom-designed test that was developed by a test publisher for use exclusively by that state. Only one state does not require an assessment; ten states require two assessments. Nine of those that require two assessments employ both a state-developed and an off-the-shelf commercial test.[1] In contrast, the remaining state in this group employs an off-the-shelf and a customized commercial test. It should be kept in mind that this is a report of the current state of affairs in state mandated reading assessment. Two states that currently use commericalcommercial, off-the-shelf assessments report that they are developing their own assessments; another reports that it will soon begin using a customized commercial test.

We found a fair amount of consistency in the ways that states report their reading test results. The great majority of states (44, or 88%) report individual student scores. Four states (8%) do not report individual student scores; however, they do report classroom, building, or district scores. While 46 states (92%) indicated that reading levels/scores were specifically reported, two others (4%) noted that their assessment was specifically developed to assess overall language arts ability, and that they had had purposely left the category score broad to reflect the complexity of the language arts.

---

1. A few additional states give off-the-shelf, nationally-normed assessments to a sample of students, and another includes a few nationally-normed items in their assessment. None of these are included in the count of states using commercial assessments.

Table 3: Types of Assessment Used for Elementary Reading, Fall 1998

Type	Percentage (*N*)*
State-Developed Assessment	60% (31)
Commercial Assessment	51% (26)
Stanford 9	14% (7)
Terra Nova	14% (7)
ITBS	14% (7)
CAT	2% (1)
Criterion-Referenced Exam	4% (2)
Choice of Several Tests	6% (3)
Customized Commercial	8% (4)
None	2% (1)

*Several states use more than one type of assessment

Alignment

States appear very committed to aligning reading and language arts standards and assessments. Only three states (6%) acknowledged a lack of alignment and foresaw a continuing lack of alignment between their standards and their assessment system. Due to various restraints, ranging from budgetary concerns to a lack of legislative or public support, these states did not anticipate progress toward alignment in the near future.

In general, the states reported using a variety of methods to determine alignment. Table 4 indicates the various approaches employed in relation to the types of assessments used. It is interesting to note that the dominant method for demonstrating alignment between standards and state-developed assessments was the state-led study. In contrast, the method used most often to determine alignment between standards and commercial assessments was the publisher-led study. Interestingly, over half of the states that had developed tests reported alignment as a result of sequential development—that is, the assessments were developed to address adopted standards. Only three of the 26 states that used commercial assessments reported that their assessments and standards were developed sequentially. In at least one of these cases, the state selected an assessment first, then adopted standards that were aligned with that assessment.

The two states that reported that the alignment of their state-developed assessments and standards had not been established also gave commercial tests. They reported, that these tests were aligned. Similarly, three of the six states that reported that their commercial assessments were not aligned gave state-developed tests that they reported as being aligned. At first glance, then, it appears that many states that gave more than one type of assessment

relied more heavily on one type or the other to judge achievement of their standards.

Table 4: Percentage and Number of States with Each Type of Assessment by Approach to Alignment*

Type of Assessment (*N*)	Sequential Development	Expert Review	Publisher-Led Study	State-Led Study	Not Established
State-Developed (31)**	52% (16)	45% (14)	6% (2)	77% (24)	6% (2)
Commercial (26)	12% (3)	15% (4)	35% (9)	27% (7)	23% (6)
Customized Commercial (4)	50% (2)	50% (2)	75% (3)	75% (3)	0% (0)

*Many states use multiple types of assessments and approaches to alignment. These are the methods used for the assessments given in the 1998–99 school year.

**Because of incomplete data, one state that uses a state-developed assessment is not included.

Summary of State Survey Results

The results of our state survey of standards and assessments indicates that by the end of 1999, 49 (98%) states will have their standards in place, and 40 states (78%) will have implemented their assessments. This increase in the number of states adopting standards and assessments appears to correspond with the changes in Title I requirements that were enacted in 1994.

Our survey indicated that state standards and assessments could be characterized by a great deal of variability in the way the domain was parsed with regard to reading and the other language arts. It also revealed that the majority of states organized both their standards and testing programs around grade-level clusters reflecting primary/elementary, intermediate/middle school, and high school levels. Although the states used a variety of types of reading/language arts assessments (e.g., state-developed and commercial), almost all states did some direct assessment of reading, apart from the other areas of the language arts. Over half of the states used some type of state-developed assessments, and many states used more than one type.

Finally, the survey revealed that the large majority of states considered their reading standards and assessments to be aligned. States used a variety of methods to determine alignment, with the state-led study being the most popular method for state-developed tests, and the publisher-led study being the most common method for the customized commercial assessments.

IV. Results of State Case Studies

State A

Officials in State A are very conscious of the importance of aligning their standards and assessments. In 1995, after the state legislature passed a law mandating that norm-referenced tests be given in kindergarten through eleventh grade, a series of committees reviewed the available commercial assessments and selected one, based in large part on their evaluations of content, administration, and norming. The form of the selected assessment is embargoed for the state. In 1997, the state approved standards at every grade level based on the categories of skills addressed in the assessment. In reading and writing, the objectives were modeled on those published with the norm-referenced test and standards developed by the National Assessment of Educational Progress. The testing company was supportive of the results of this process. Alignment of the norm-referenced test to the standards, then, is achieved through sequential development of the test and then the standards. A state-led study has verified this alignment.

For the purposes of this study, we analyzed the two third-grade reading standards (reading comprehension and reading vocabulary) and the norm-referenced test that was used to assess third-grade students' attainment of these standards. Although we did not consider it in this analysis, State A also mandates a state-developed writing test in the fourth, seventh, and eleventh grades. This test was developed after the standards, and alignment was established through sequential development.

Characterization of Standards

State A's elementary language arts standards include listening, speaking, reading comprehension, reading vocabulary, writing, spelling, language, study skills, and technology. The standards that target reading, reading comprehension and reading vocabulary have the same labels as the sections of the norm-referenced test that target reading. Reading comprehension is explained by 20 objectives; reading vocabulary is explained by five. Most of these objectives are concise, specific, and easy to understand. Examples include "use context clues to determine meaning," "draw conclusions about a sequence of activities in an announcement or advertisement," and "recognize the correct meaning of a word with multiple meanings when presented in text." The objective we judged to be most cognitively complex is "identify theme, main idea, and author's purpose in a selection when it is not explicitly stated." We judged it to be in the third level of cognitive complexity, because demonstrating the attainment of this objective could include identifying abstract themes across a text.

Characterization of Assessment

The reading section of the third-grade assessment consists of 84 multiple-choice items, of which 30 are devoted to reading vocabulary and 54 to reading comprehension. The reading comprehension section includes nine passages: four fiction and five non-fiction. The authors of the four fictional passages are indicated. The nine reading selections range in length from 110 to 302 words, with the average length being 197. The reading vocabulary section includes 18 items which ask students to choose the correct defini-

tion of a word, six items which require students to use context clues in isolated sentences to identify the correct definition of a word with multiple meanings, and six items which ask then to choose the correct definition of a word with the aid of sentence context clues.

At present, State A determines students' level of proficiency in meeting the standards by considering their percentile scores on the norm-referenced test. Those who score at or above the fiftieth percentile are judged to achieve satisfactory performance. Those below the fiftieth percentile are judged to have made unsatisfactory progress toward the achievement of the state standards. School accountability for student performance begins at the third grade, and is based in large part on a basic skills score derived from the norm-referenced test. In addition to a readiness test in kindergarten, students in kindergarten, first, and second grade take the norm-referenced test, but these scores are used for diagnostic purposes only.

Alignment Criteria

Range of knowledge correspondence and balance of representation. The second column in Table A.1 indicates that 80% of the total objectives in reading explain the reading comprehension standard, and that State A seems to place more emphasis on this standard than on reading vocabulary. The assessment, however, as indicated by the final column, places more emphasis on vocabulary than do the objectives. It is interesting to note that the final column totals 100%: a result of the fact that each item hit either one standard or the other, but not both. This reflects the method which State A used to ensure alignment—writing standards based on the test.

Table A.1: Range of Knowledge Correspondence and Balance of Representation Among Standards, Objectives, and Assessment Items for State A

Reading Standard	Number and % of Objectives for Each Standard Related to Total Number of Objectives	Number and % of Assessment Items for Each Standard Related to Total Number of Items
1. Reading Comprehension	20/25 (80%)	54/84 (64%)
2. Reading Vocabulary	5/25 (20%)	30/84 (36%)

Coverage. By our judgment, over 64% (or approximately two-thirds) of the objectives in reading are addressed by the assessment (see Table A.2). Of the seven unassessed reading comprehension objectives, three are difficult to assess in an on-demand, paper-and-pencil test: e.g., "read literary works by national and international authors to include, but not limited to: legends, folktales, and non-fiction," "chooses and responds to a variety of reading material for pleasure and information," and "experience content through imagery (visualization)." One of the reading vocabulary objectives—"given a variety of reading material, increase the number of recognized words presented in text"—faces the same constraint.

According to our analysis, the test emphasizes the assessment of two objectives in particular. A comprehension objective—"identify explicitly stated information including, but not limited to: story elements (e.g., setting, characters, plot), a set of directions, and functional reading (e.g. invitations, bulletins)"—is addressed by 26 items. A vocabulary objective—"recognize synonyms, antonyms, homonyms, and homophones for identified vocabulary words presented in isolation or within a group of words"—is assessed by 18 items. Three other objectives (one in reading comprehension and two

BEST COPY AVAILABLE

in vocabulary) are assessed by six items each. Of the remaining eleven comprehension objectives that are represented in the test, three are assessed by only one item and eight are assessed by two or three items.

Table A.2: Assessment Item Coverage of Standards and Objectives for State A

Reading Standard	Number and % of Objectives for Each Standard Represented by at Least One Assessment Item
1. Reading Comprehension	13/20 (65%)
2. Reading Vocabulary	3/5 (60%)
TOTAL	16/25 (64%)

Depth of knowledge consistency. Table A.3 reveals that the cognitive level of the reading comprehension objectives is, on average, slightly higher than that of the assessment items with which they are associated. Approximately two-thirds of the reading comprehension objectives are rated at Level 2 or above, while only one-third of the reading comprehension items are rated at Level 2. None are rated above Level 2. The cognitive levels of the reading vocabulary objectives and assessment items are, on the other hand, highly aligned at the lower end of the cognitive complexity continuum. Overall, the average cognitive level of the objectives is 1.56, and the level of the assessment is 1.21—the lowest of any of the state assessments evaluated in this study.

Table A.3: Depth of Knowledge Consistency for State A

Reading Standard	Cognitive Level of Objectives Within Each Standard	Cognitive Level of Items Related to Each Standard
1. Reading Comprehension	7/20 (35%) at Level 1 12/20 (60%) at Level 2 1/20 (5%) at Level 3 0 at Level 4	36/54 (67%) at Level 1 18/54 (33%) at Level 2 0 at Level 3 0 at Level 4
2. Reading Vocabulary	5/5 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	30/30 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
3. Average for each level	12/25 (48%) at Level 1 12/25(48%) at Level 2 1/25 (4%) at Level 3 0 at Level 4	66/84 (79%) at Level 1 18/84 (21%) at Level 2 0 at Level 3 0 at Level 4

Structure of knowledge comparability. State A's standards and objectives can be best characterized by the mastery model of curriculum. For the most part, the objectives read as a list of discrete skills that can be understood independently of the reader. These objectives include "identify explicitly stated information," "determine sequence," "recognize characteristics of a fictional and non-fictional story." Just two of the 25 reading objectives reflect a curriculum model other than mastery. These two—"make predictions based on prior knowledge and story information" and "chooses and responds to a variety of reading material for pleasure and information"—could arguably be located in a cognitive model because they address students' application of prior knowledge and personal response to create meaning from text. Even these two objectives, however, include elements—

BEST COPY AVAILABLE

making predictions from story information, and choosing texts for information—that best reflect a mastery model. None of the objectives approach a social-constructivist perspective.

The norm-referenced test administered by State A is also best characterized by the mastery model. The exam emphasizes skills such as word identification, identification of word definitions in isolated sentences, and verbatim recall of information from passages. Although the exam includes some items that require students to infer, the information needed to make the inferences is explicit in the text. The exam does not require students to relate reading to their own experiences, or to participate actively in the construction of meaning. State A's standards and assessment are fully grounded in a mastery model of curriculum and are, therefore, well-aligned in this criteria.

Issues from the Analysis of State A

Our analysis finds that state A's standards, objectives, and assessment items have a relatively high degree of alignment. State A raises the following issues regarding alignment:

1. Because skills like those in State A's vocabulary objectives and items are relatively easy to assess in a paper-and-pencil test, they may be over-represented in some assessments.

2. It may be necessary to consider the number of items assessing each standard or objective, especially in relation to the criteria for proficiency. In State A, many objectives are assessed by only a few items. This, combined with the fact that the minimum criteria for satisfactory performance is that students reach the 50% percentile, may mean that a student can be judged to have performed satisfactorily on given a standard if he or she correctly answered a small number of items that addressed only a portion of the objectives within that standard.

3. Although State A's norm-referenced test is generally aligned to its standards, the standards and objectives (and especially the assessment items) are, on average, of a low level of cognitive complexity. We argue that this assessment—the only assessment of reading for which schools are held accountable—does not meet the Title I requirement that assessment measures include "measures that assess higher order thinking skills and understanding" (Part A, Subpart 1, Section 1111, (b), (3), (E)).

4. The very fact that State A judges students' proficiency according to their percentile scores brings the state's alignment into question. The notion of standards-based assessment requires that each student's academic attainment is judged against the criteria established by the standards. Judging proficiency by comparing students to their peers runs counter to the spirit of Title I. Setting this issue aside, we wonder how many items a student must answer correctly to be judged proficient. Practically speaking, the bar may not be set very high.

State B

State B approved its current assessment system in 1990, and first implemented it in the 1990-91 school year. A state-developed reading comprehen-

sion test is administered in grades 3 through 8, and a writing assessment is given in grades 4, 8, and 10. The standards developed by State B for English language arts performance were approved in 1997. For our analysis we examined the reading standards and objectives for grade 4, and the reading portion of the grade 4 state assessment. State B considers its assessment and standards to be aligned, and reports that the methods of establishing alignment were sequential development (with the assessment preceding the standards), a state-led study, and an expert review.

Characterization of Standards

State B has developed standards and objectives for each individual grade level, K-12. In addition to the reading standards and objectives which we examined as part of our analysis, their framework document includes standards and objectives for Listening/Speaking (with subcategories in purpose, culture, audiences, communication) and Writing (with subcategories in purposes, penmanship/capitalization/punctuation, spelling, grammar/usage, writing processes, evaluation, inquiry/research). Each subcategory contains a standard followed by several objectives. This state's standards document includes over 100 objectives for English language arts in the fourth grade. As stated in the introduction to this document, the standards and objectives are designed to meet state legislation's goal that "the students in the public education system will demonstrate exemplary performance in the reading and writing of English language." It is also the state's goal that all children will be reading at grade level by the end of third grade.

Characterization of Assessment

The State B reading test is a reading comprehension test. It includes six passages, each followed by between four and eight multiple-choice questions. The passages include both fiction and nonfiction. The passages range in length from approximately 100 words to approximately 450 words. The passing rate for the exam was determined by the State Board of Education. In order to pass, students must answer 70% of the multiple-choice items correctly. This is a high-stakes assessment, with a school's overall passing rate (along with percentages of drop-outs and attendance levels) determining its designation as either exemplary, recognized, acceptable, or low-performing.

Alignment Criteria

Range of knowledge correspondence and balance of representation. Table B.1 demonstrates the assessment's emphasis on reading comprehension, as opposed to the other eight standards. Thirty-eight of 40 (or 95%) of the assessment items reflect one or more of the reading comprehension objectives. However, the reading comprehension objectives make up only 22% of the total number of objectives. The standards for Text Structures/Literary Concepts and Inquiry/Research, which account for 19% and 15% of the total objectives respectively, are represented by only 8% and 10% of assessment items. In contrast, some standards account for a higher percentage of assessment items than total number of objectives. For example, Word Recognition accounts for just 5% of the total objectives, but 13% of the assessment items relate to those objectives. Vocabulary development accounts for just 9% of the total objectives, but 18% of the assessment items.

Overall, it seems that this test's focus is comprehension, and that therefore it does not accurately measure the other standards. Tables B.1 and Table B.2 together show that although 95% of the assessment relates to the reading comprehension objectives, only 67% of the objectives in this standard are covered by at least one assessment item. Eight of the 12 reading comprehen-

sion objectives are covered by 38 of 40 assessment items. This leaves 4 objectives in reading comprehension unaddressed in the assessment. Of 63 total times that an assessment item "hit" a standard, 39 of those were in reading comprehension. The three standards that are not represented at all in the assessment—culture, fluency and variety of texts—are difficult to assess with a paper-and-pencil test. In short, if we focus only on comprehension, then the test appears to align well to the objectives. If on the other hand the entire set of standards and objectives is considered, then there is far less alignment.

Table B.1: Range of Knowledge Correspondence and Balance of Representation Among Standards, Objectives, and Assessment Items for State B

Standard	Number and % of objectives for each standard related to total number of objectives	Number and % of assessment items for each standard related to total number of items
1. Reading/Word Recognition: The student uses a variety of word recognition strategies.	3/54 (5%)	5/40 (13%)
2. Reading/Fluency: The student reads with fluency and understanding texts at appropriate difficulty levels.	6/54 (11%)	0/40 (0%)
3. Reading/Variety of Texts: The student reads widely for different purposes in varied sources.	3/54 (5%)	0/40 (0%)
4. Reading/Vocabulary Development: The student acquires an extensive vocabulary through reading and systematic word study.	5/54 (9%)	7/40 (18%)
5. Reading/Comprehension: The student comprehends selections using a variety of strategies.	12/54 (22%)	38/40 (95%)
6. Reading/Literary Response: The student expresses and supports responses to various types of texts.	4/54 (7%)	2/40 (5%)
7. Reading Text Structures/Literary Concepts: The student analyzes the characteristics of various types of texts (genres).	10/54 (19%)	3/40 (8%)
8. Reading/Inquiry/Research: The student inquires and conducts research using a variety of sources.	8/54 (15%)	4/40 (10%)
9. Reading/Culture: The student reads to increase knowledge of own culture, culture of others, common elements of culture	3/54 (5%)	0/40 (0%)

Coverage. Again, the State B assessment covers the reading comprehension objectives to a much greater extent than it covers the objectives in the other standards. As shown in Table B.2, after comprehension objectives, the objectives reflected next most often in the assessment are those under the Vocabulary Development standard. Two of the five objectives in this standard, or 40%, are represented in the assessment. Six of the nine standards are represented by at least one assessment item, but only 30% of the total number of objectives is represented by the assessment. Although six of the nine standards are represented in some way in the assessment, this relatively

high number is somewhat misleading since just one or two objectives are addressed in every standard except comprehension and only between two and seven assessment items represent any standard other than comprehension.

Table B.2: Assessment Item Coverage of Standards and Objectives for State B

STANDARD	NUMBER AND % OF OBJECTIVES FOR EACH STANDARD REPRESENTED BY AT LEAST ONE ASSESSMENT ITEM
1. Reading/Word Recognition:The student uses a variety of word recognition strategies.	1/3 (33%)
2. Reading/Fluency:The student reads with fluency and understanding texts at appropriate difficulty levels.	0/6 (0%)
3. Reading/Variety of Texts:The student reads widely for different purposes in varied sources.	0/3 (0%)
4. Reading/ Vocabulary Development:The student acquires an extensive vocabulary through reading and systematic word study.	2/5 (40%)
5. Reading/Comprehension:The student comprehends selections using a variety of strategies.	8/12 (67%)
6. Reading/Literary Response:The student expresses and supports responses to various types of texts.	1/4 (25%)
7. Reading/Text Structures/Literary Concepts:The student analyzes the characteristics of various types of texts (genres).	2/10 (20%)
8. Reading/Inquiry/Research:The student inquires and conducts research using a variety of sources.	2/8 (25%)
9. Reading/Culture:The student reads to increase knowledge of own culture, culture of others, common elements of culture	0/3 (0%)
Total	16/54 (30%)

Depth of knowledge consistency. Tables B.3 and B.4 present information about the comparative levels of cognitive complexity between standards/objectives and assessment items. All of the objectives under the reading comprehension standard were rated as a 2 (83%) or a 3 (17%).This indicates that all of those objectives require more than literal recall of information, and also require a certain amount of intersentence or across-passage analysis, inference, summarizing, and theme identification. However, many of the items that represent those objectives in the assessment were given a rating of 1 for cognitive complexity (40%). Forty-eight percent of the items that mapped onto the comprehension standard were rated a 2, and just 10% of items were rated as a 3.Although the standard and objectives do not emphasize literal recall or identification, many of the comprehension assessment items only require that superficial level of analysis.

Similarly, Standard 8 includes no objectives rated at a level 1, while 25% of its objectives were rated a 2, 63% were rated a 3, and 13% were rated a 4. However, 100% of the items that relate to that standard were rated as a 1. The standard requires a level of cognitive complexity that is clearly not embedded in the assessment. In general, the assessment items tend to be less cognitively complex than the objectives. Indeed, the averages across objectives (2.1) and items (1.7) suggest that the cognitive complexity of standards and assessment items is not closely aligned (see Table B.4).

Table B.3: Depth of Knowledge Consistency for State B

Standard	Cognitive Level of Objectives Within Each Standard	Cognitive Level of Items Related to Each Standard
1. Reading/Word Recognition:The student uses a variety of word recognition strategies.	3/3 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 5/5 (100%) at Level 2 0 at Level 3 0 at Level 4
2. Reading/Fluency:The student reads with fluency and understanding texts at appropriate difficulty levels.	5/6 (83%) at Level 1 1/6 (17%) at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
3. Reading/Variety of Texts:The student reads widely for different purposes in varied sources.	2/3 (66%) at Level 1 1/3 (33%) at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
4. Reading/ Vocabulary Development:The student acquires an extensive vocabulary through reading and systematic word study.	0 at Level 1 5/5 (100%) at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 7/7 (100%) at Level 2 0 at Level 3 0 at Level 4
5. Reading/Comprehension:The student comprehends selections using a variety of strategies.	0 at Level 1 10/12 (83%) at Level 2 2/12 (17%) at Level 3 0 at Level 4	16/39 (41%) at Level 1 19/39 (49%) at Level 2 4/39 (10%) at Level 3 0 at Level 4
6. Reading/Literary Response:The student expresses and supports responses to various types of texts.	0 at Level 1 0 at Level 2 4/4 (100%) at Level 3 0 at Level 4	0 Level 1 1/2 (50%) at Level 2 1/2 (50%) at Level 3 0 at Level 4
7. Reading/Text Structures/Literary Concepts:The student analyzes the characteristics of various types of texts (genres).	0 at Level 1 6/10 (60%) at Level 2 4/10 (40%) at Level 3 0 at Level 4	0 at Level 1 2/6 (33%) at Level 2 4/6 (67%) at Level 3 0 at Level 4
8. Reading/Inquiry/Research:The student inquires and conducts research using a variety of sources.	0 at Level 1 2/8 (25%) at Level 2 5/8 (63%) at Level 3 1/8 (13%) at Level 4	4/4 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
9. Reading/Culture:The student reads to increase knowledge of own culture, culture of others, common elements of culture.	0 at Level 1 0 at Level 2 3/3 (100%) at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
Total by Level	10/54 (19%) at Level 1 25/54 (46%) at Level 2 18/54 (33%) at Level 3 1/54 (2%) at Level 4	16/63 (25%) at Level 1 20/63 (32%) at Level 2 4/63 (6%) at Level 3 0 at Level 4

Table B.4: Cognitive Levels of Total Objectives and Test Items

Standards	Items
10/54 (19%) at Level 1 25/54 (46%) at Level 2 18/54 (33%) at Level 3 1/54 (2%) at Level 4	16/40 (40%) at Level 1 20/40 (50%) at Level 2 4/40 (10%) at Level 3 0 at Level 4
Average: 2.13	Average: 1.7

BEST COPY AVAILABLE

Structure of knowledge comparability. State B's standards appear to best reflect a cognitive model of curriculum. Their standards include reference to prior knowledge and the application of the readers' experiences in determining meaning of text. To cite one example, a fourth-grade objective reads: "support responses by referring to relevant aspects of text and own experiences." The objectives emphasize inquiry, requiring students to ask questions of text and investigate questions across sources. For example, another fourth-grade objective reads: "offer observations, make connections, react, speculate, interpret, raise questions, in response to text." These objectives are typical of several across-grade levels that emphasize students' role as active participants in the meaning-making process. Although several objectives do emphasize basic skills, and a few require students to examine their own experiences in relation to those of others, the document as a whole does not reflect either a mastery or a social-constructivist perspective.

The State B assessment reflects both a mastery and a cognitivist model. Many of the assessment items ask students to recall specific information from the text, thus testing the students' abilities to decode and locate information. Items that do not require students to do more than locate information in a passage reflect a mastery model. Other items, however, require that the student infer information from the story. Several questions require students to choose the main idea or the best summary of a passage. These questions require students to participate more actively in the meaning-making process. Although they do not explicitly require respondents to apply prior knowledge, many of them do so implicitly. A very few items seem to reach beyond a mastery model. However, much of the test reflects a mastery model, and to be fully aligned a test would need to more often reflect the cognitive model embedded in the standards and objectives.

Issues from Analysis of State B

Our analysis finds that state B's standards, objectives, and assessment items have a relatively low degree of alignment. State B raises the following issues regarding alignment:

1. The exam focuses on a specific area of reading, such as reading comprehension, to the exclusion of other areas included in the state's framework. State B's assessment is partially aligned to its reading comprehension standard, but the assessment does not adequately address the other standards. Given that this state's assessment was most likely designed as a measure of reading comprehension, it seems reasonable to focus our discussion of alignment on the relationship between the exam and the one standard that it most fully addresses. However, the fact that the majority of the objectives in the state's framework are unaddressed by the assessment has implications for instruction and student achievement. Teachers are most likely to focus attention on those areas of the curriculum for which they are held accountable. In this case, the reading comprehension standard may receive more instructional time than the eight other standards, because comprehension is clearly the focus of the assessment.

2. The cognitive complexity of the standards and assessment suggest that the assessment tasks are of lower cognitive complexity than the objectives. Although none of the reading comprehension objectives was rated as a cognitive Level 1, 41% of the items related to that standard were rated at a Level 1. This misalignment could potentially affect the cognitive complex-

ity of the reading comprehension tasks emphasized in the state's classrooms.

3. Although the state reports that alignment was established in part through sequential development, with the assessment having preceded the standards, it does not appear that the standards were written in light of this assessment. It is laudable that the state constructed a set of relatively demanding, detailed standards and objectives that address many areas within reading. A state-developed test written in light of these objectives would be a positive step toward ensuring that these objectives are emphasized in classrooms, and toward a determination of whether students are meeting the goals set forth by the framework.

State C

Motivated largely by the Title I legislation, State C began developing a student assessment system in the early 1990s. By 1997, they had developed language arts standards and implemented a testing system. State testing consists of a state-developed reading comprehension test given in third grade, as well as a norm-referenced test given in grades 4, 8, and 10. The reading comprehension test is a "student-level" test, and although the results are published and distributed throughout the state, the state has not analyzed the test's alignment with the standards. The core of the testing program is the norm-referenced test. Its alignment to the standards was established through a workshop led by the test's publisher. The study found that 55% of the standards overlapped with the norm-referenced test. Ninety-eight percent of the norm-referenced test items are aligned with that 55% of the standards.

Although State C has not included its reading comprehension test in an alignment study, we believe that a consideration of its alignment is critical. If instruction is to be consciously aligned with challenging standards, then all components of the assessment and accountability system must be aligned with those standards. We argue that the results of any test whose results are disseminated throughout the state is, de facto, part of the assessment system. As a result, our analysis included State C's fourth-grade reading standards, the state-developed reading comprehension test given in the third grade, and the reading portion of the norm-referenced test given in the fourth grade.

Characterization of Standards

State C has developed standards at grades 4, 8, and 10, the grades tested with the norm-referenced test. When developing its standards, State C drew from a variety of sources, including in-state studies and numerous externally-written standards. State C has four reading and literature standards: "students will use reading strategies to achieve their purposes in reading," "read, interpret, and critically analyze literature," "read and discuss literary and nonliterary texts in order to understand human experience," and "read to acquire information." Each standard is explained in more detail by objectives. Many of the objectives are very long and incorporate many skills and processes. For example, "use a variety of strategies and word recognition skills, including rereading, finding context clues, applying their knowledge of letter-sound relationships, and analyzing word structures" and "summarize ideas drawn from stories, identifying cause-and-effect relationships, interpreting

events and ideas, and connecting different works to each other and to real-life experiences." This made it difficult to determine whether or not a particular assessment item was assessing a given objective.

It is also more difficult for us to evaluate alignment when the standards and objectives overlap. For example, six of Standard 1's eight objectives identify reading strategies; two do not. Objective 4 includes "comprehend reading by... establishing purpose...." Objective 8 is "identify a purpose for reading such as..." We also had difficulty evaluating the extent to which assessment items related to overly general standards/objectives. For example, Standard 4 is "read to acquire information," which could refer to almost anything in a reading comprehension assessment. In this case, we inferred on the basis of the accompanying objectives that this standard was intended to refer to interaction with non-fiction text.

Characterization of Assessments

State-developed reading comprehension test. The state-developed reading comprehension test consists of 104 multiple-choice items. Sixteen of these items assess students' prior knowledge of concepts and the vocabulary needed for easy comprehension of two passages. Because these items are included for purposes of school and district analysis of results and do not contribute to student scores, we disregarded them in our assessment of alignment. The test is comprised of three passages: two fiction (1126 and 1543 words) and one non-fiction (629 words). The average passage length is 1099 words.

State C reports students' performance on the test in four categories: advanced, proficient, basic, and minimal. The cutoff scores for these levels are determined by the state. Although the exact numbers vary slightly by year, students who answer approximately 79% or more of the items correctly achieve the proficient level. Students who fail to show proficiency may be considered for remedial reading services.

Norm-referenced test. The reading language arts sections of the norm-referenced test which is given to students at the end of fourth grade consist of 64 multiple-choice items. Nineteen of these items, however, are not included in our present analysis because they assess either grammar or writing, as opposed to reading. A few of the 45 analyzed items are in fact writing or grammar items, but are included in the analysis anyway, because they are related to reading passages and could serve as evidence of student attainment of a reading standard. The analyzed portion of the test contains six passages: four fiction and two non-fiction. These passages range in length from 133 words to 423 words; the average passage length is 297 words. State C uses the cut scores provided by the publishing company to determine whether student attainment of standards is minimal, basic, or advanced. Pattern scores, as opposed to raw scores, are used to determine students' proficiency levels. As a result, we can say that students who achieve the proficient level answered approximately 67% of the items correctly. The actual number of correct answers, however, will vary according to students' patterns of answering simple and difficult questions.

Alignment Criteria

Range of knowledge correspondence and balance of representation. Table C.1 indicates that over 40% of the total objectives explain Standard 1, "students will use reading strategies to achieve their purposes in reading." The rest of the objectives are fairly evenly distributed among the other three

standards. The vast majority of items in both the state-developed reading comprehension test and the norm-referenced test assess Standard 1. More than half of the total items assess Standard 2.

Most of the items in both tests address only a few objectives—"comprehend reading by using strategies such as activating prior knowledge, establishing purpose, self-correcting..."(Standard 1, Objective 4), "recognize and recall elements and details of story structure... in order to reflect on meaning" (Standard 2, Objective 1), and "extend the literal meaning of a text by making inferences, and evaluate the significance and validity of texts (Standard 2, Objective 4). This finding reflects our decision rules. All but 15 of the items on the reading comprehension test were read by the students and therefore require comprehension, and so assessed Standard 1, Objective 4. Only seven items, all on the reading comprehension test, do not assess Standard 1. These seven do not require understanding of the preceding passage in the test, and their question prompts and possible responses are read to the students by the test administrator. Similarly, any item that required the recall of details in the story was judged to assess Standard 2, Objective 1, despite the fact that few items fully addressed that objective. Finally, any item that required an inference from the text was judged to assess Standard 2, Objective 4, despite the fact that no items required students to evaluate the significance and validity of texts. Our assessment of balance, then, is generous.

Coverage. Table C.2 reveals that approximately two-thirds of the standards and objectives are covered by at least one assessment item. This table also indicates that the state-developed test does not add much coverage beyond that provided by the norm-referenced test. Given the inattention to alignment of the reading comprehension test, it is not surprising that only about 40% of the objectives are addressed by that test. The norm-referenced test, on the other hand, addresses the majority of objectives in 3 of the 4 standards.

Four of the seven objectives that are not covered by either assessment seem fairly difficult to assess in a paper-and-pencil test—"demonstrate phonemic awareness by using letter/sound relationships as aids to pronouncing and understanding unfamiliar words and text," (Standard 1, Objective 3) " read aloud with age-appropriate fluency, accuracy, and expression," (Standard 1, Objective 5) "identify a purpose for reading, such as gaining information, learning about a viewpoint, and appreciating literature," (Standard 1, Objective 8) and "identify a topic of interest then seek information by investigating available text resources" (Standard 4, Objective 2). The other three objectives that are not covered by either assessment seem relatively easy to address in an on-demand, paper-and-pencil test. "Identify and use organizational features of texts, such as headings, paragraphs, and format, to improve understanding" (Standard 1, Objective 7) can be at least partially assessed in a paper-and-pencil test. The other two perhaps could be more fully addressed. They are "distinguish fiction from nonfiction, realistic fiction from fantasy, biography from autobiography, and poetry from prose" (Standard 3, Objective 3) and "summarize key details of informational texts, connecting new information to prior knowledge" (Standard 4, Objective 1).

BEST COPY AVAILABLE

Table C.1: Range of Knowledge Correspondence and Balance of Representation Among Standards, Objectives, and Assessment Items for State C

Reading Standard	Number and % of Objectives for Each Standard Related to Total Number of Objectives	State-Developed: Number and % of Items for Each Standard Related to Total Number of Items	Norm-Referenced: Number and % of Items for Each Standard Related to Total Number of Items	Combined Total: Number and % of Items for Each Standard Related to Total Number of Items
1. Use reading strategies to achieve purposes	9/21 (43%)*	69/77 (90%)	45/45 (100%)	114/122 (93%)
2. Read, interpret, and critically analyze literature	5/21 (24%)	*44/77 (57%)	19/45 (42%)	63/122 (52%)
3. Read and discuss literary and nonliterary texts in order to understand the human experience	4/21 (19%)	3/77 (4%)	10/45 (22%)	13/122 (11%)
4. Read to acquire information	3/21 (14%)*	22/77 (29%)	6/45 (13%)	28/122 (23%)

* Standard is counted as an objective. Although no objective under the standard was assessed, the standard itself was assessed by at least one item.

Table C.2: Assessment Item Coverage of Standards and Objectives for State C

Reading Standard	State-Developed: Number and % of Objectives for Each Standard Represented by at least one Assessment Item	Norm-Referenced: Number and % of Objectives for Each Standard Represented by At Least One Assessment Item	Overall: Number and % of Objectives for Each Standard Represented by At Least One Assessment item
1. Use reading strategies to achieve purposes	4/9 (44%)*	5/9 (56%)	5/9 (56%)
2. Read, interpret, and critically analyze literature	3/5 (60%)*	4/5 (80%)	5/5 (100%)
3. Read and discuss literary and nonliterary texts in order to understand the human experience	1/4 (25%)	3/4 (75%)	3/4 (75%)
4. Read to acquire information	1/3 (33%)*	1/3 (33%)	1/3 (33%)
TOTAL	9/21 (43%)	13/21 (62%)	14/21 (66%)*

* Standard is counted as an objective. Although no objective under the standard was assessed, the standard itself was assessed by at least one item.

Depth of knowledge consistency. Table C.3 reveals that the objectives' level of cognitive complexity is slightly higher than that of the test items. The level of cognitive complexity appears to be higher for the items assessing Standard 1, which requires students to "use reading strategies to achieve purposes." Conversely, the level of cognitive complexity required by the objectives appears to be higher than that required by the items assessing those objectives for Standard 2 ("read, interpret, and critically analyze literature) or Standard 3 ("read and discuss literary and nonliterary texts in order to understand the human experience). The only substantial difference in the tests appears to be that the items assessing Standard 3 are at higher levels of cognitive complexity on the state-developed reading comprehension test than they are on the norm-referenced test. Table C.4 indicates that the overall levels of cognitive complexity are almost identical for the items in the reading comprehension test and the items in the norm-referenced test. The average cognitive level of the objectives is higher, but not dramatically so.

BEST COPY AVAILABLE

Table C.3: Depth of Knowledge Consistency for State C

Reading Standard	Cognitive Level of Objectives Within Each Standard	State-Developed: Cognitive Level of Items Related to Each Standard	Norm-Referenced: Cognitive Level of Items to Each Standard	Overall Total: Cognitive Level of Items Related to Each Standard
1. Use reading strategies to achieve purposes*	7/9 (78%) at Level 1 1/9 (13%) at Level 2 0 at Level 3 0 at Level 4	41/69 (59%) at Level 1 27/69 (39%) at Level 2 1/69 (1%) at Level 3 0 at Level 4	28/48 (42%) at Level 1 19/48 (40%) at Level 2 1/48 (2%) at Level 3 0 at Level 4	69/117 (59%) at Level 1 46/117 (39%) at Level 2 2/117 (2%) at Level 3 0 at Level 4
2. Read, interpret, and critically analyze literature*	0 at Level 1 1/5 (20%) at Level 2 3/5 (60%) at Level 3 1/5 (20%) at Level 4	9/44 (20%) at Level 1 34/44 (77%) at Level 2 1/44 (2%) at Level 3 0 at Level 4	5/22 (23%) at Level 1 16/22 (73%) at Level 2 1/22 (.5%) at Level 3 0 at Level 4	14/66 (21%) at Level 1 50/66 (76%) at Level 2 2/66 (3%) at Level 3 0 at Level 4
3. Read and discuss literary and nonliterary texts in order to understand the human experience	0 at Level 1 3/4 (75%) at Level 2 1/4 (25%) at Level 3 0 at Level 4	0 at Level 1 2/3 (67%) at Level 2 1/3 (33%) at Level 3 0 at Level 4	2/10 (20%) at Level 1 8/10 (80%) at Level 2 0 at Level 3 0 at Level 4	2/13 (15%) at Level 1 10/13 (77%) at Level 2 1/13 (8%) at Level 3 0 at Level 4
4. Read to acquire information*	1/3 (33%) at Level 1 2/3 (67%) at Level 2 0 at Level 3 0 at Level 4	19/22 (86%) at Level 1 3/22 (14%) at Level 2 0 at Level 3 0 at Level 4	5/6 (83%) at Level 1 1/6 (17%) at Level 2 0 at Level 3 0 at Level 4	24/28 (86%) at Level 1 4/28 (14%) at Level 2 0 at Level 3 0 at Level 4
TOTAL BY COGNITIVE LEVEL	8/21 (38%) at Level 1 8/21 (38%) at Level 2 4/21 (19%) at Level 3 1/21 (5%) at Level 4	69/138 (50%) at Level 1 66/138 (48%) at Level 2 3/138 (2%) at Level 3 0 at Level 4	40/86 (47%) at Level 1 44/86 (51%) at Level 2 2/86 (2%) at Level 3 0 at Level 4	109/224 (49%) at Level 1 110/224 (49%) at Level 2 5/224 (2%) at Level 3 0 at Level 4

* Standard is counted as an objective because neither objective under the standard was assessed but the standard itself was assessed by at least one item.

Table C.4: Cognitive Levels of Total Objectives and Test Items

	Objectives	Reading Comprehension Test	Norm-Referenced Test	Total Test Items
	8/21 (38%) at Level 1 8/21 (38%) at Level 2 4/21 (19%) at Level 3 1/21 (5%) at Level 4*	41/77 (53%) at Level 1 35/77 (45%) at Level 2 1/77 (1%) at Level 3 0 at Level 4	25/45 (56%) at Level 1 19/45 (42%) at Level 2 1/45 (2%) at Level 3 0 at Level 4	66/122 (54%) at Level 1 54/122 (44%) at Level 2 2/122 (2%) at Level 3 0 at Level 4
Averages	1.84	1.48	1.47	1.48

*The total of 21 reflects that three of the standards are counted as objectives. Although no item specifically addresses any objective under these standards, the standard itself is assessed by at least one item.

Structure of knowledge comparability. The standards and objectives in State C best reflect a cognitive model of curriculum; however, they also include several objectives from a mastery model. Many of the objectives are best characterized by the cognitive model because they emphasize the reader's active role in constructing meaning from text. For example, "summarize key details of informational texts, connecting new information to prior knowledge," and "Extend literal meaning of text by making inferences, and evaluate the significance and validity of texts" each include elements—one the application of prior knowledge, and the other evaluation—that require students to draw on personal experience and previous encounters with texts. This state's standards also include several objectives that reflect a mastery model, but overall their standards are best understood as coming from a cognitive model.

The state-developed reading comprehension test reflects a mastery model of curriculum, and emphasizes verbatim recall of information from passages. Although the exam includes some items that require students to infer, the information needed to make the inference is always explicitly stated in the text. For example, for the item "Which word best describes the [character in the passage]?" only one of the four possible answers makes sense in light of information in the passage. The exam does not require students to relate reading to their own experiences or participate actively in the construction of meaning.

The norm-referenced exam also predominantly reflects a mastery model of curriculum. Like the items in the state-developed reading comprehension test, the majority of reading items in the norm-referenced exam emphasizes verbatim recall from text and inferences based on information found in the passage. A very few items approach a cognitive model by allowing students to make connections across passages. For example, one item asks students to write a paragraph comparing the experiences of a character in a cartoon to a character in a poem. This item emphasizes the readers' role in constructing meaning, as they apply their knowledge of one text in order to understand another. State C's standards and objectives reflect both a mastery and a cognitive model, but favoring the cognitive. The opposite is true of the assessments, one of which includes a very few items reflecting a cognitive model. The vast majority of items from both tests reflects a mastery model. State C's standards and assessments are, therefore, only partially aligned in the area of structure of knowledge.

Issues from Analysis of State C

Our analysis finds that state C's assessment system and reading standards have a relatively high degree of alignment. State C's case raises the following issues regarding alignment:

1. Complex, list-like objectives may result in artificially high alignment given the methodology employed in our analysis. Objectives are so lengthy and complex that many items address only a small fraction of the intended skills or processes. As a result, the assessment system looks more aligned than it truly is. For example, we judged that many items address "summarize ideas drawn from stories, identifying cause-and-effect relationships, interpreting events and ideas, and connecting different works to each other and to real-life experiences." No item or combination of items, however, fully addresses this objective. In other cases, a single item might touch on portions of many objectives. Finally, some items address a standard itself, but no objective within that standard seemed a match.

2. Dramatic differences in the length of the assessments' passages are not well reflected in our evaluation of the items' cognitive complexity. The average length of the passages in the state-developed reading comprehension test (approximately 1,100 words) is considerably longer than the average length of passages in the norm-referenced test (approximately 300 words). Logically, the items associated with the longer passages are more difficult to answer. More text must be retained, summarized, or reviewed. Our ratings of cognitive complexity, however, do not reflect the differences associated with text length.

3. To ensure alignment, states must think in terms of assessment systems, as opposed to tests. Administering a supplemental test without regard to alignment may not change the level of alignment achieved. In this case, the reading comprehension test does not cover any objectives that are not covered by the norm-referenced test. Furthermore, the two tests show similar cognitive complexity. In terms of accountability, the reading comprehension test may not warrant the additional effort and expense. Although two tests require different

percentages of correct answers in order for students to achieve proficiency (fewer items must be answered correctly on the norm-referenced test), the actual scores provide similar information. If the norm-referenced test can provide the sorts of diagnostic information that the state gleans from the reading comprehension test, then it may be reasonable to give only the norm-referenced test, or, perhaps better, to move the norm-referenced test to third grade and suspend the reading comprehension test.

State D

State D's standards and assessments were adopted and implemented in 1996. In elementary language arts this system includes two exams—a nationally developed criterion-referenced test (CRT) administered in the fourth grade, and a state-developed second-grade individualized oral reading assessment (IRA). To establish the alignment between standards and assessment in State D, we examined both of these tests and the reading standards for the elementary grades (preK–4th grade). In addition to the grade four administration, the nationally developed CRT is administered in grades eight and eleven. The state believes that these exams are in alignment with their curriculum framework. They established alignment by constructing the standards first and then developing exams in light of those standards. Although the CRT was not developed by the state alone, the state was involved in the development process, and their own standards were written in part to reflect the standards on which the exam is based.

Characterization of Standards

State D's English language arts standards are: Communication (which includes Reading and Writing), and Arts/Language and Literature (which includes Critical Response, Literature and Media, the English Language, Non-Native Language, Artistic Process and Elements, Forms, and Techniques in the Arts). The standards are written for clusters of grades: pre-K to 4, 5 through 8, and 9 through 12. The Communication standards include a total of 52 objectives at the preK to fourth-grade levels.

The purpose of State D's standards, as described in their document, is to improve student learning. They expect the standards to be used in three ways:

1. To provide a structure from which standards-based district, school and classroom curriculum can be developed, organized, implemented and assessed.

2. To provide the basis for the development of a state, local, and classroom comprehensive assessment system.

3. To make explicit what may be included in statewide assessments of student learning.

The developers of the standards view them as practical, helpful reference points for the development of curriculum and assessment, and as targets for achievement.

Characterization of Assessments

Nationally developed CRT. The nationally developed CRT includes 30 items scored for reading or reading and writing, and 11 items scored only for writing. In our analysis, we excluded the items scored only for writing. The reading questions include passages ranging in length from one paragraph to two pages, followed by multiple-choice and short-answer questions, and one question that requires extended writing. The passages include fiction, non-fiction and informational text.

The exam is scored against performance levels that are referenced to a set of national standards, which strongly resemble the standards in this state. The state reports that the test developers have determined the performance levels by which students are judged to have achieved the standards. The state defines students

with a rating of "proficient" as having achieved the standard. Scores are reported for individual students, classrooms, schools, and districts. The scores are reported by standard, and include advice about how a student can work to improve their performance on the standard.

Individualized reading assessment (IRA). The IRA includes a read-aloud and an oral retelling. In the read-aloud portion of the assessment the administrator takes a record of oral reading, or keeps running records of the child's oral reading after determining the level at which the child comfortably reads. In the oral retelling, the child is asked to retell a story that they have just read. The child's score is based on the number of the story's major elements which are included in the retelling. Score reports indicate whether a child has exceeded, achieved, approached, or scored below the standard. The standard is based on a 90% reading accuracy for grade-level text, and an acceptable comprehension rate as determined by a scoring rubric.

Alignment Criteria

Range of knowledge correspondence and balance of representation. Table D.1 suggests that State D's objectives and assessments do not cover the same range of topics, and that the topics that are similar across these two groups are not given the same level of attention. However, in State D's framework it appears that the number of objectives included in each standard does not reflect the relative importance of the standards, but rather the amount of detail deemed necessary to convey information about what children should know and be able to do in that particular area of reading. For example, 46% of the objectives in the reading framework for this state are located in Standard 1—a far greater percentage than for Standard 3. The objectives for Standard 1, however, are a list of reading strategies that the developers of the framework deemed important for students in the early grades to know and use. These include "sounds, syllables, and letter patterns," syntax," "meaning in context," "a range of cueing systems," "self-correcting," "questioning," and "prior knowledge." This list of skills requires Standard 1 to include more total numbers of objectives, but does not necessarily imply that Standard 1 be given more curricular attention than the other standards.

Although Standard 3 accounts for the smallest proportion of the total number of objectives (20%), 100% of the assessment items in the national exam relate to this Standard. It appears that the national exam is focused on reading comprehension, and does not attend to the other areas of emphasis in State D's standards. In contrast, the two items that comprise the IRA—oral reading and retelling—relate to three of the four standards. The oral reading portion of the test relates to Standards 1 and 2, and the retelling item relates to Standard 3. The apparent imbalance between the attention given to Standard 2 in the objectives and in the IRA is an artifact of the decision to consider reading accuracy as a separate standard, with no accompanying objectives.

Although Standard 4 is not addressed in either assessment, the nature of the objectives in this standard makes them difficult to measure in a statewide accountability system. These objectives relate to dispositions toward reading, encouraging students to spend time reading across texts and genres. The objectives measure both the breadth and depth of children's reading across a school year. For example, students will "read at least 25 books a year...," "read at least three different kinds (genres) of printed materials and at least five different writers," "read primary and secondary sources," and "read at least four books about one issue or subject...."

Coverage. Table D.2 indicates that all of the items on the national CRT mapped to the objectives in Standard 3, which addresses reading for meaning or reading comprehension. The objectives in this standard state that students should be able to "comprehend grade-appropriate materials," "analyze and interpret features of a variety of types of text," and "make connections among various parts of a text, among several texts, and between texts and other experiences in and out of school." Each of these three objectives is addressed in at least one item on the national exam. The reading portion of the exam seems to be focused exclusively on reading comprehension, and does not address other objectives in the State D framework.

BEST COPY AVAILABLE

Table D.1: Range of Knowledge Correspondence and Balance of Representation Among Standards, Objectives, and Assessment Items for State D

Reading Standard	Number and % of Objectives for Each Standard Related to Total Number of Objectives	Number and % of Assessment Items for Each Standard Related to Total Number of Items		
		National CRT	State IRA	Overall
1. Students use a variety of strategies to help them read.	7/15 (46%)	0	1/2 (50%)	1/2 (50%)
2. Students read grade-appropriate material with 90%+ accuracy, in a way that makes meaning clear.	1/15 (6%)*	0	1/2 (50%)	1/2 (50%)
3. Students read for meaning, demonstrating both initial understanding and personal response to what is read.	3/15 (20%)	30/30 (100%)	1/2 (50%)	31/32 (97%)
4. Students comprehend and respond to a range of media, images, and text.	4/15 (26%)	0	0	0

*Standard counted as objective because this standard contains no objectives

Table D.2: Assessment Item Coverage of Standards and Objectives for State D

Reading Standard	Proportion of Objectives Under Each Standard Represented by At Least One Assessment Item		
	National CRT	State IRA	Overall
1. Students use a variety of strategies to help them read.	0	7/7 (100%)	7/7 (100%)
2. Students read grade-appropriate material with 90%+ accuracy, in a way that makes meaning clear.	0	1/1 (100%)*	1/1 (100%)
3. Students read for meaning, demonstrating both initial understanding and personal response to what is read.	3/3 (100%)	1/3 (33%)	3/3 (100%)
4. Students comprehend and respond to a range of media, images, and text.	0	0	0
TOTAL	3/14 (21%)	9/14 (64%)	11/14 (79%)

*Standard counted as objective because this standard contains no objectives

Standards 1 and 2, which are not explicitly addressed in the national exam, are covered in the state IRA. The records of oral reading provide teachers with an opportunity to observe the strategies that children use to decode text, and thus to assess the objectives included in Standard 1. In addition, the oral reading portion of the test fully addresses Standard 2 by allowing teachers to assess the level of accuracy at which children are reading. The retelling portion of the IRA also addresses the first objective in Standard 3—students' ability to comprehend grade-appropriate text. The state-published materials accompanying the IRA explicitly state that the test is designed to address Standards 2 and 3 of the state curriculum framework. We found that the test does address both of those standards, but that it also measures all of the objectives under Standard 1. Between the national exam and the state IRA, State D's assessment system addresses all of the objectives except those in Standard 4, which are not easily measured in either the paper-and-pencil test or an oral reading/retelling test.

Depth of knowledge consistency. Table D.3 reveals that almost half of the objectives in State D's framework appear under Standard 1. In addition, all of the objectives under Standard 1 are rated at a Level 1 in cognitive complexity, in contrast to the Standard 3 objectives, two out of the three of which are rated at a Level 3. Both the objectives in Standard 3 and the assessment items relating to them require more complex thought, on average, than the objectives and assessment items for Standards 1 or 2. Overall, 80% of the objectives in the standards are rated as Level 1, while just 30% of the assessment items are rated as Level 1.

The analysis of level of cognitive complexity presented in Table D.4 suggests that the level of cognitive complexity of the assessment items is somewhat higher than that of the objectives. Contrary to this apparent trend, however, the only objectives that count in the national reading exam are those in Standard 3, which showanshow an average cognitive complexity of 2.3, reflecting a higher level of complexity than the assessment which measures them (see Table D.3). At the same time, it is worth noting that the national exam includes a higher proportion of items at a cognitive level of 2 or above than did the assessments in our other focus states. The majority of items in the national exam are rated at a Level 2 for cognitive complexity, requiring more than verbatim recall of information from the text. In addition, the test includes items that require students to organize their thoughts and to infer the author's purpose for writing the passage. With just 20% of the items rated at a Level 1, the national exam requires students to do more than recall literal information from text, which is in keeping with two of the three reading comprehension objectives.

The IRA's level of cognitive complexity aligns well with the objectives that it measures. The oral reading portion of the test is rated at a Level 1 because it requires only that the student decode a text, without analysis or interpretation. The retelling portion of the test is rated at a Level 2 because it requires students to organize the information thahtthat they have read and present that information in an oral summary. This item requires students to integrate information from the story and share that information in their own words, much like an assessment item from the national exam that requires the student to rewrite the story in another form.

BEST COPY AVAILABLE

Table D.3: Depth of Knowledge Consistency for State D

Reading Standard	Cognitive Level of Objectives Within Each Standard	Cognitive Level of Items for Each Standard: National CRT	Cognitive Level of Items for Each Standard: State IRA	Overall Cognitive Level of Items for Each Standard
1. Students use a variety of strategies to help them read.	7/7 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	8/8 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	8/8 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
2. Students read grade-appropriate material with 90%+ accuracy, in a way that makes meaning clear.	1/1 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	1/1 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	1/1 (100%) at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
3. Students read for meaning, demonstrating both initial understanding and personal response to what is read.	1/3 (33%) at Level 1 0 at Level 2 2/3 (66%) at Level 3 0 at Level 4	9/46 (20%) at Level 1 19/46 (41%) at Level 2 2/46 (4%) at Level 3 0 at Level 4	0 at Level 1 1/1 (100%) at Level 2 0 at Level 3 0 at Level 4	9/47 (19%) at Level 1 20/47 (43%) at Level 2 2/47 (4%) at Level 3 0 at Level 4
4. Students comprehend and respond to a range of media, images, and text.	3/4 (75%) at Level 1 1/4 (25%) at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4	0 at Level 1 0 at Level 2 0 at Level 3 0 at Level 4
Totals	12/15 (80%) at Level 1 1/15 (6%) at Level 2 2/15 (13%) at Level 3 0 at Level 4	9/46 (20%) at Level 1 19/46 (41%) at Level 2 2/46 (4%) at Level 3 0 at Level 4	9/10 (90%) at Level 1 1/10 (10%) at Level 2 0 at Level 3 0 at Level 4	18/56 (32%) at Level 1 20/56 (36%) at Level 2 2/56 (4%) at Level 3 0 at Level 4

Table D.4: Cognitive Levels of Total Objectives and Test Items

	Objectives	National CRT	State IRA	Total Test Items
	12/15 (80%) at Level 1 1/15 (6%) at Level 2 2/15 (13%) at Level 3 0 at Level 4	9/30 (30%) at Level 1 19/30 (63%) at Level 2 2/30 (6%) at Level 3 0 at Level 4	1/2 (50%) at Level 1 1/2 (50%) at Level 2 0 at Level 3 0 at Level 4	10/32 at Level 1 20/32 at Level 2 2/32 at Level 3 0 at Level 4
Averages	1.33	1.77	1.5	1.75

Structure of knowledge comparability. Three of State D's four standards reflect a mastery model of curriculum, while Standard 3 reflects a cognitive model. The objectives in Standards 1, 2, and 4 emphasize discrete skills and/or quantifiable tasks. Standard 1 addresses the skills and strategies necessary to read and comprehend text, but does not ask the student to construct meaning from her/his own experience or apply the information in text to other meaningful contexts. Standard 2 addresses a student's ability to read text fluently, and Standard 4 states the numbers and types of texts and genres that a student is required to read in each school year. In contrast, Standard 3 includes objectives that stress the analysis and interpretation of texts and connections between texts and experience—activities that emphasize the readers' role in creating meaning through interactions with text. The standards do not include any objectives that reflect a social-constructivist model of curriculum.

The state IRA measures second-grade students' ability to read fluently with literal comprehension, and thus reflects a mastery model of curriculum. The purpose of the assessment is not to measure student's ability to interpret text, or to relate text to their own experiences or to other texts they have read. The fourth grade national reading exam also primarily reflects a mastery model. None of the reading items require students to apply their prior knowledge of a topic, relate their reading to their own experiences or to other texts, or work with others to construct responses; instead, they require students to recall information verbatim from the passage or infer from information available in the passage. State D's standards and assessments predominantly reflect a mastery model of curriculum and are, therefore, well-aligned in structure of knowledge.

Issues from Analysis of State D

Our analysis finds that State D's assessment system and reading standards have a moderate degree of alignment. State D's case raises the following issues regarding alignment:

1. All but one of the standards are assessed by the state's accountability system. Some standards are difficult to address in statewide accountability systems, particularly those, such as Standard 4, which address dispositions toward reading and students' reading across texts and genres.

2. The cognitive complexity of the assessments can be higher than that of the standards and objectives.

3. State D is the only state of the four in our analysis to include a state-developed individualized reading assessment. This exam allows State D to assess areas of the reading framework that a paper-and-pencil test cannot measure and, therefore, increases their degree of alignment.

Summary and Discussion of State Case Studies

A summary of information regarding Range of Knowledge Correspondence, Balance of Representation, Coverage, and Depth of Knowledge Consistency appears in Table 5. Based on these criteria, along with the Structure of Knowledge analysis, it appears that the standards, objectives, and assessments are reasonably well aligned in States A and C, moderately well aligned in State D, and relatively poorly aligned in State B. The alignment status of these states appears to be related to their history of statewide reading curriculum and assessment, as well as to their unique approaches to the issue of alignment. Following are some of the most important distinctions:

- State A achieved alignment by relying heavily on the norm-referenced test which they planned to use to generate their standards and objectives. Although there appears to be a relatively high level of alignment in State A, it is important to note that many of their objectives are assessed by only one or two items on the norm-referenced test. This, in combination with a performance standard of the fiftieth percentile, suggests that State A students could be considered proficient without demonstrating competence on all state objectives covered by the assessment.

- State B has a long history of heavily skills-oriented curriculum and assessment in reading. The relatively low level of alignment observed in State B appears to stem from the development of more sophisticated standards and objectives without a corresponding revision of the state test. The low level of alignment is apparent in all five of the criteria used in this analysis.

- State C is fairly new to reform, and uses a combination of norm-referenced and state-developed exams. The relatively high level of alignment observed in this state appears to be related to the fact that their standards and objectives are so general that they apply to almost any and every assessment item. This conclusion may also be an artifact of the type of document analysis employed in our study.

- State D has a long history as a reforming state, and uses a nationally-developed, criterion-referenced exam in combination with an individualized oral reading assessment. State D's moderate level of alignment stems from two characteristics of their standards and objectives—their breadth, and their relatively low level of cognitive complexity. The breadth of the standards and objectives makes them difficult to assess—particularly those focused on dispositions. The relatively low level of the standards' and objectives' cognitive complexity does not align well with the nationally-developed criterion-referenced exam that they selected, which has the highest level of cognitive complexity of any of the assessments we evaluated. On the positive side, State D's inclusion of an individualized oral reading assessment as part of their assessment system makes their alignment stronger than it would be with the national exam alone.

Table 5: Summary of Range, Balance, Coverage, and Depth of Knowledge by State

	Number of Items Assessing Each Standard Over Total Number of Standards (Range and Balance)				Percentage of Objectives Assessed in Those Standards Addressed by 1 or More Items (Coverage)				Similarity of Cognitive Levels of Objectives and Items (Depth of Knowledge)	
	0	1–4	5–8	8+	<25%	25–50%	50–75%	>75%	Average CL of objectives	Average CL of items
State A				2/2			2/2		1.56	1.21
State B	3/9	3/9	2/9	1/9	1/6	4/6	1/6		2.13	1.7
State C				4/4		1/4	1/4	2/4	1.84	1.48
State D	1/4			3/4		1/3		2/3	1.33	1.75

* State D's individualized reading test is equated to at least eight items.

Looking across the states, we clearly cannot assume that the level of cognitive complexity represented by a set of standards will be reflected in the corresponding assessment, or vice versa. This point is illustrated by the contrast between States B and D. In State B, the cognitive level of the assessment items appears to be lower than that of the standards and objectives, with 40% of the items rated at a Level 1, and 80% of the standards and objectives rated at a Level 2 or above. In contrast, State D's national CRT exam appears to be more cognitively complex than its standards and objectives. Almost 70% of the assessment items are rated at a Level 2 or above, while 80% of the standards and objectives are rated at a Level 1.

Looking across all four focus states, we can also see that the use of more than one assessment may or may not improve alignment. For example, State D is the only state of the four that has developed an individualized oral reading assessment to supplement its norm-referenced test. This decision allows State D to assess objectives not addressed by the other states. In contrast, State C administers two tests that measure similar objectives.

In general, we were struck by the diversity among the focus states on the issue of alignment. Our impression, after working with this information, is that the alignment data do not—and, perhaps, cannot—tell the entire alignment story for any given state. Rather, alignment data must be supplemented with information gathered from supplementary documents and personal interviews, which describe the development and selection of standards and assessments, and each state's "take" on the issue of alignment.

BEST COPY AVAILABLE

V. Implications

What did we learn from this analysis of alignment between state standards and assessments in elementary reading? We learned that the issue of alignment is more complex than it appears at first glance. First, alignment means different things in different states. Most states indicated that their standards and assessments were aligned. If our case studies are at all representative of other states, however, we could expect low to moderate alignment in a quarter to a half of the states claiming alignment. Our study also revealed that states approach alignment in a variety of ways. Some spend considerable energy ensuring that their assessments are aligned with their standards; some do not. For some, alignment is an integral part of the annual development or approval of tests. For others, it is a one-time study. Some states recognize that their assessment system is not aligned with their standards, but do not have the financial resources to write the tests which they feel they need.

We also learned that the method of determining alignment is likely to have an impact on the outcome of the analysis. We grappled with a number of issues as we conducted this document analysis, including: the assumption that standards are all written at the same level of specificity, and that their importance can be determined by the number of objectives associated with each standard; questions about the role of performance standards in evaluating alignment; and the discovery that high levels of alignment do not equate to high-quality standards and assessments. The method we used to establish alignment appears to be most applicable when the test items are easily identifiable and weighted evenly in the scoring of the assessment. It may be that more authentic evaluations of student performance, like the individualized assessment in State D, do not lend themselves as well to our counting procedures.

Neither the state's methods of determining scores nor their cut scores for proficient, advanced, and partially proficient students are systematically taken into account in our method for determining alignment. In most of the assessments that we evaluated, it was fairly simple to determine whether or not an item contributed to the reading score. When this was unclear, we perhaps inflated the level of alignment by including all items that we believed could contribute to a reading score. This allowed us to understand the ways in which information from the assessment could be best used to demonstrate student attainment of standards. However, it would also make sense to try to understand how the state actually does determine a reading score from the assessment. Similarly, information on how each state determines proficiency levels would allow us to more fully analyze how challenging the assessment truly is.

Further, we began to wonder about the differences between evaluating alignment of reading/language arts and doing the same for other subject areas, such as mathematics and science. For example, our beliefs about reading/language arts as holistic processes made us more comfortable reporting our analyses at the level of the standards, as opposed to the objectives. We did not want to give the impression that because this type of analysis reveals high levels of alignment between discrete sets of standards and assessment items that this is an accurate reflection of our understanding of the domain of English language arts. We wondered, for example, if perhaps the domain would be better reflected if we focused on passages and their accompanying items as the unit of analysis within reading assessments. We also reflected on the extent to which levels of cognitive complexity were comparable across subject areas, since it is necessary to assume mastery of a whole set of basic word recognition and comprehension skills before one achieves at even the lowest levels of cognitive complexity in reading.

It seems clear to us that there is much yet to be learned about both the concept of alignment and the methods used for alignment analyses—particularly in the area of English language arts. We should tread lightly when drawing conclusions and making generalizations about alignmentuntilalignment until we have more experience in this important area.

References

Mathematical Sciences Education Board. (1993). *Measuring what counts. A conceptual guide for mathematics assessment.* Washington, DC: National Academy Press.

Payzant, T. W. & Levin, J. (1993). Improving America's schools for children in greatest need. In J. F. Jennings (Ed.), *National issues in education: The past is prologue.* Bloomington, IN: Phi Delta Kappa.

Rothman, R. (December 5, 1997). *Summary of first meeting of the Committee on Title I Testing and Assessment.* Washington, DC: National Reading Council.

Rothman, R. & Elmore, R. (1997). *Standards-based reform: The "theory of action."* Washington, DC. Unpublished paper prepared for the Committee on Title I Testing and Assessment.

Smith, M. L. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona student assessment program.* Los Angeles: UCLA, National Center for Research on Evaluation, Standards, and Student Testing.

Webb, N. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Washington, DC: Council of Chief State School Officers.

Appendix A: Interview Protocol

State _____

Person Interviewed _____

Title of Position Held _____

Phone Number _____

Date of Interview _____

Consent Statement: My name is _____ and I am calling from the University of Michigan. We are working on a study commissioned by the National Research Council's Committee on Title I Testing and Assessment and would like to ask you a few questions regarding the alignment between standards and assessments in your state. I will need approximately 15 minutes of your time. Your involvement is completely voluntary and you can choose to stop this interview at any time. Your identity will be kept confidential and your name will not appear in any report related to this study. May I begin the interview?

1. Could you describe your assessment system for Elementary Language Arts including all elements that schools and/or districts must report to the state?

I've put you in the (state-developed, unmodified publisher designed, customized publisher designed) category. Would you agree?

Elements of Assessment System

	Test Only	Test and other Items
State-Developed		
Publisher Designed (e.g., Stanford 9, Terra Nova)		
Custom Publisher Designed		
Other Items (e.g. portfolios)		

If custom-designed, is it a parallel form of a published test with restricted use or have you customized the test to meet the objectives in your state?

2. Given these five categories, how would you characterize the method or methods through which your state has established alignment (Sequential Development, Expert Review, Publisher-led Study, State-led Study, Not Established)? (Could be more than one.)

States' Reports of Method Used to Establish Alignment Between Standards and Assessment System

	Sequential Development	Expert Review	Publisher-led Study	State-led Study	Not Established
State Developed					
Publisher Designed					
Custom Publisher Designed					
Other Items (e.g., portfolios)					

BEST COPY AVAILABLE

3. Is reading assessed independently of other areas of language arts such as writing? (Is there a separate test for reading or is it integrated, for example, with a writing test?)

If yes—In what categories are scores reported?

If no—So, reading is assessed in conjunction with other areas of language arts such as writing. In what categories are scores reported?

We also obtained the information needed to fill in the following chart either from state web pages, documents, or phone interviews.

State

**Process by which assessment system was approved		
**Yr Stand apprvedapproved		
**Yr Assmt system approved		
Assessment Type		
Supplementary tests/elements		
**Yr Assmt first used (not pilot)		
**Grades Assessed		
**Grades of Standards		
Is assessment aligned to standards?		
How is alignment established?		
Is reading assessed independently of writing?		
**In what categories do you report LA scores?		
**Indiv Scores reported?		
**How do you determine scores for advanced, proficient, novice		
**Contact		

Appendix B: Sample Coding Sheet

NOTE: The coding sheet has been split to fit the format of this report. Each page contains several complete rows of data.

		CL.ITEMS →	1	1	1	1	1	1	1	1	1	1	1	1	1
Obj	STATE A-3rd grade	Item# →	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Reading Comprehension	CL.Obj														
15	read literary works by natl and internatl authors to incl, but not lmt to, legends, folk-tales, nonfict	1														
16	id explicitly stated info incl,butnotlmt2,story elemts, set of directions, functional read-ing	1														

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	2	2	1	2	1	2	1	1	2	1	1	2
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
																1			1		1	1		1	1			1		.1	2	1			1		

2	2	1	1	2	1	2	1	1	1	2	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1.226190476 (AVG. CL ITEMS)
23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	
																																# of hits
																																0
			1		1		1	1	1					1	1			1		1	1	1		1			1		1		1	26

BEST COPY AVAILABLE

		CL ITEMS →	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Obj	STATE A-3rd grade	Item# →	1	2	3	4	5	6	7	8	9	10	11	12	13	14
17	determine sequence (eg events in a story, set of dirs, missing item)	2														
18	use context clues to determine word meaning	1														
19	recognize characteristics of a fictional and non-fict story	2														
20	draw concls regarding char (eg feelings, moods, traits, motives, pt of view)	2														
21	draw concl about a sequence of activities in an announcemt or ad	2														
22	draw concl regarding the identity of certain objs when spec details are given	2														

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	2	2	1	2	1	2	1	1	2	1	1	2
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
																																				1	
																																	1				
																	2																	2				
																		1																				

2	2	1	1	2	1	2	1	1	1	2	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1.226190476 (AVG. CL ITEMS)
23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	# of hits
																							1									2
		1																												1		2
2																																1
																																3
																									1							1
																																1

		CL ITEMS →	1	1	1	1	1	1	1	1	1	1	1	1	1
Obj	STATE A-3rd grade	Item# →	1	2	3	4	5	6	7	8	9	10	11	12	13	14
23	compare & contrast in order to draw concl regarding a story	2														
24	make predictions based on prior know and story info	2														
25	determine an appropriate title for a reading selection or a story	2														
26	id theme, main idea, and author's purpose in a selection when it is not explicitly stated	3														
27	paraphrase, summarize, compose ?s, & make inferences about material read	2														
28	recognize other resources where addnl info on a specific topic can be located	1														

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	2	1	2	1	2	1	1	2	1	1	2
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
																				2																	
																							2				2									
																													2							2

2	2	1	1	2	1	2	1	1	1	2	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1.226190476 (AVG. CL ITEMS)
23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	# of hits
																																0
																2																2
																	2															3
	2										2	2																1				6
						2				2									1													3
													1													1						2

BEST COPY AVAILABLE

BEST COPY AVAILABLE

		CL Items →	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Obj	STATE A-3rd grade	Item# →	1	2	3	4	5	6	7	8	9	10	11	12	13	14
29	id poss sources of functional info (eg where would such info probably appear	1														
30	id cause and effect related to a given event (eg what happened and why)	2														
31	make generalizations regarding story elements	2														
32	recog figurative language (eg similes, metaphors, and idioms)	1														
33	chooses and responds to a var of reading material for pleasure and info	2														
34	experience content through imagery (visualizing)	1														

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	2	2	1	2	1	2	1	1	2	1	1	2
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
																										2											

2	2	1	1	2	1	2	1	1	1	2	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1.226190476 (AVG. CL ITEMS)
23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	# of hits
				2																												2
																																0
																																0
																																0
																																0
																																0

		CL ITEMS →	1	1	1	1	1	1	1	1	1	1	1	1	1
Obj	STATE A-3rd grade	Item# →	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	reading vocabulary															
35	recog synonyms, antonyms, homonyms, & homophones for ided vocab wrds presented in isolation or within a group of words	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
36	recog the correct meaning of a wrd w/mult meanings when presented in text	1														
37	apply structural analysis and context clues to decode and encode words	1														
38	id and use content area vocab	1														
39	given a var of reading mat, increase the # of recognized words presented in text	1														
	average CL of objectives	1.56														

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	2	2	1	2	1	2	1	1	2	1	1	2
15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1							1	1	1	1	1	1																						
				1	1	1	1	1	1																												

2	2	1	1	2	1	2	1	1	1	2	2	2	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1.226190476 (AVG. CL ITEMS)
23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	# of hits
																																24
																																6
																																0
																																0
																																0

BEST COPY AVAILABLE

APPENDIX C: Cognitive Levels Rubric

Depth of Knowledge

The cognitive level of each objective and each item is judged using the following scale:

1 *Level 1* requires students to receive or recite facts or to use simple skills and abilities. Oral reading that does not include analysis of the text as well as basic comprehension of a text is included. Questions require only a shallow understanding of text presented and often consist of verbatim recall from text or simple understanding of a single phrase.

2 *Level 2* requires some mental processing; it requires both comprehension and subsequent processing of text or portions of text. Intersentence analysis or inference is required. Some important concepts are covered but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed.

3 Deep knowledge becomes more of a focus at *Level 3*. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Standards and items at this level involve reasoning and planning. Students must be able to support their thinking. Questions may involve abstract theme identification or inference across an entire passage. Questions may also involve more superficial connections between texts.

4 Higher order thinking is central and knowledge is deep at *Level 4*. The standard or the item at this level will probably be an extended activity, with extended time provided. Students take information from at least one passage and are asked to apply this information to a new task. They may also be asked to develop hypotheses and perform complex analyses of the connections among texts.

APPENDIX D: Structure of Knowledge Rubric

Structure of Knowledge

The structure of knowledge of each assessment and each set of standards is evaluated against the following models of language arts curriculum.

Mastery	The *mastery model* is closely associated with competency assessment, instructional objectives, and basic skills learning. Skills are viewed as discrete, separable entities that can be understood apart from and independent of the learner/reader. Successful progression through sets of skills is taken as evidence of satisfactory achievement in reading. Standards and assessment items that fit this model focus on student performance of isolated skills, such as the ability to punctuate sentences correctly.
Cognitive	The *cognitive model* emphasizes the relationship between the knowledge that a reader brings to a text and the reader's ability to comprehend that text. Readers are not being viewed as passive receptacles of knowledge from text, but are rather seen as active participants in meaning-making, processing new knowledge by drawing on already existing knowledge gleaned from personal experiences and previous encounters with text. Standards and assessment items fitting this model require students to draw on prior knowledge to construct meaning from text and/or relate a passage read to something in their experience.
Social-constructivist	Rather than viewing knowledge as an individual construction, a *social-constructivist position* views knowledge as collective and contingent upon human beings' interactions with and shared beliefs about the world. Language is recognized as a social and cultural construction and, therefore, literacy practices cannot be viewed apart from the social and cultural contexts in which they are learned and practiced. Standards and assessment items that fit this model connect students' language learning to social context. For example, a standard may require that students engage in literacy activities that engage an issue in their community. An assessment item may require that students work with others to construct an answer to writing prompt.

About CIERA

The Center for the Improvement of Early Reading Achievement (CIERA) is the national center for research on early reading and represents a consortium of educators in five universities (University of Michigan, University of Virginia, and Michigan State University with University of Southern California and University of Minnesota), teacher educators, teachers, publishers of texts, tests, and technology, professional organizations, and schools and school districts across the United States. CIERA is supported under the Educational Research and Development Centers Program, PR/Award Number R305R70004, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

Mission. CIERA's mission is to improve the reading achievement of America's children by generating and disseminating theoretical, empirical, and practical solutions to persistent problems in the learning and teaching of beginning reading.

CIERA Research Model

The model that underlies CIERA's efforts acknowledges many influences on children's reading acquisition. The multiple influences on children's early reading acquisition can be represented in three successive layers, each yielding an area of inquiry of the CIERA scope of work. These three areas of inquiry each present a set of persistent problems in the learning and teaching of beginning reading:

CIERA INQUIRY 1
Readers and Texts

Characteristics of readers and texts and their relationship to early reading achievement. What are the characteristics of readers and texts that have the greatest influence on early success in reading? How can children's existing knowledge and classroom environments enhance the factors that make for success?

CIERA INQUIRY 2
Home and School

Home and school effects on early reading achievment. How do the contexts of homes, communities, classrooms, and schools support high levels of reading achievement among primary-level children? How can these contexts be enhanced to ensure high levels of reading achievement for all children?

CIERA INQUIRY 3
Policy and Profession

Policy and professional effects on early reading achievement. How can new teachers be initiated into the profession and experienced teachers be provided with the knowledge and dispositions to teach young children to read well? How do policies at all levels support or detract from providing all children with access to high levels of reading instruction?

www.ciera.org

CIERA is a collaboration of
University of Michigan
University of Virginia
Michigan State University
with
University of Minnesota
University of Southern California

University of Michigan School of Education
610 E University Av, Rm 1600 SEB
Ann Arbor, MI 48109-1259
734.647.6940 voice
734.615.4858 fax
ciera@umich.edu

www.ciera.org

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
X	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").